

Simulation-Based Analysis of Effects of *Vrn* and *Ppd* Loci on Flowering in Wheat

Jeffrey W. White,* Markus Herndl, L. A. Hunt, Thomas S. Payne, and Gerrit Hoogenboom

ABSTRACT

Cereal production is strongly influenced by flowering date. Wheat (*Triticum aestivum* L.) models simulate days to flower by assuming that development is modified by vernalization and photoperiodism. Cultivar differences are parameterized by vernalization requirement, photoperiod sensitivity, and earliness per se. The parameters are usually estimated by comparing simulations with field observations but appear estimable from genetic information. For wheat, the *Vrn* and *Ppd* loci, which affect vernalization and photoperiodism, were logical candidates for estimating parameters in the model CSM-Cropsim-CERES. Two parameters were estimated conventionally and then re-estimated with linear effects of *Vrn* and *Ppd*. Flowering data were obtained for 29 cultivars from international nurseries and divided into calibration (14 locations) and evaluation (34 locations) sets. Simulations with a generic cultivar explained 95% of variation in flowering for calibration data (10 d RMSE) and 89% for evaluation data (10 d RMSE), indicating the large effect of environment. Nonetheless, for the calibration data, the gene-based model explained 29% of remaining variation, and the conventional model, 54%. For the evaluation data, the gene-based model explained 17% of remaining variation, and the conventional model, 27%. Gene-based prediction of wheat phenology appears feasible, but more extensive genetic characterization of cultivars is needed.

J.W. White, USDA-ARS, Arid Land Agricultural Research Center, 21881 N. Cardon Ln., Maricopa, AZ 85239; M. Herndl, Institute of Crop Production and Grassland Research (340), Univ. of Hohenheim, Fruwirthstr. 23, D-70599 Stuttgart, Germany; L.A. Hunt, Dep. of Plant Agriculture, Crop Science Bldg., Univ. of Guelph, Guelph, ON, Canada N1G 2W1; T.S. Payne, International Maize and Wheat Improvement Centre (CIMMYT), Ap. Postal 6-641, 06600, Mexico DF, Mexico; G. Hoogenboom, Dep. of Biological and Agricultural Engineering, College of Agricultural and Environmental Sciences, Univ. of Georgia, Griffin, GA 30223-1797. Mention of a specific product name by the U.S. Department of Agriculture does not constitute an endorsement and does not imply a recommendation over other suitable products. Received 7 June 2007. *Corresponding author (jeffrey.white@ars.usda.gov).

Abbreviations: CIMMYT, International Maize and Wheat Improvement Centre; IWIS, International Wheat Information System; IWHPN, International Winter Wheat Performance Nurseries; QTL, quantitative trait loci; SS, sum of squares.

TIMING OF FLOWERING is a critical determinant of the adaptation of a grain or seed crop to a given production environment and set of management practices. In bread wheat (*Triticum aestivum* L.), the *Vrn* and *Ppd* families of loci exert major influences on flowering by modifying the sensitivity of reproductive development to vernalization and photoperiod, respectively (Flood and Halloran, 1986; Worland, 1996; Laurie et al., 2004; van Beem et al., 2005; Iqbal et al., 2007). Although wheat researchers recognize the importance of these loci, attempts to quantitatively predict the effects of different loci under variable environment conditions are rare. A robust, quantitative methodology for interpreting interacting effects of specific loci and the environment

Published in Crop Sci. 48:678–687 (2008).

doi: 10.2135/cropsci2007.06.0318

© Crop Science Society of America

677 S. Segoe Rd., Madison, WI 53711 USA

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

should permit a much more mechanistic understanding of genotype by environment interactions for phenology, grain yield, and other economically important traits.

Process-based ecophysiological models can integrate the effects of environment, crop management, and cultivar to predict crop growth and development. Cultivar differences in phenology are usually embodied through parameters that modify sensitivity to factors such as vernalization temperatures or photoperiod regimes (e.g., Ritchie, 1991). Cultivar differences in earliness per se may also be represented through minimum durations of specific developmental phases. Often labeled “genetic coefficients” to reflect their use in specifying cultivar differences, the cultivar parameters nonetheless are estimated primarily using phenotypic data obtained from field trials (White and Hoogenboom, 1996; Baenziger et al., 2004). This process is often slow and expensive since obtaining reliable calibrations may require conducting field trials at multiple locations or over several cropping seasons. The process also is often considered prone to over-calibration for local conditions, which can inflate the apparent validity of a model, hinder detection of problematic assumptions or errors in other model inputs, and reduce the reliability of model-based predictions.

Recognizing the problems inherent in use of cultivar-specific model parameters, various researchers have pursued strategies to replace the parameters with data for genetic loci. The GeneGro model for common bean (*Phaseolus vulgaris* L.) used data for seven loci and predicted phenology, yield, and grain size as well as the parent model using conventionally estimated cultivar parameters (White and Hoogenboom, 1996; Hoogenboom et al., 1997, 2004b; Hoogenboom and White, 2003). Messina et al. (2006) successfully modeled phenology in soybean using the same approach. Efforts to use data for quantitative trait loci (QTL) instead of Mendelian loci have shown promise for modeling specific traits (e.g., Reymond et al., 2003; Nakagawa et al., 2005). Attempts to parameterize inputs for whole plant models using QTL appear somewhat less encouraging (Yin et al., 2000). The problems encountered may reflect difficulties inherent in accurately phenotyping the large numbers of lines required for QTL analyses and in identifying true QTL through appropriate statistical analysis (Edmeades et al., 2004). Recently, several papers have argued for more mechanistic approaches for integrating genomics with modeling (e.g., Minorsky, 2003; Tardieu, 2003; White and Hoogenboom, 2003; Hammer et al., 2004; Struik et al., 2005; Wollenweber et al., 2005; Yin et al., 2004; White, 2006). Actual applications of such approaches are few. Among promising examples are efforts to model gene networks for flowering at a simplified level (Welch et al., 2003).

Over 40 Mendelian loci have been described as affecting physiological traits in wheat (White, 2006). Of these,

perhaps 30 affect traits that are readily modeled. Thus, it appears that there is sufficient genetic information to initiate work on a gene-based wheat model. Experience with GeneGro (White and Hoogenboom, 1996) suggested that the linear regressions used to estimate effects of genes on model parameters could be programmed external to the model code. Given the large number of wheat models available (e.g., Grant et al., 2001; Jamieson et al., 1998; Asseng et al., 2002), it seemed desirable to develop a gene-based estimator of cultivar-specific model parameters, rather than a single gene-based wheat model (White, 2006). If successful, this would facilitate direct transfer of results to alternative wheat models.

The goal of this study was to test the practicability of using a gene-based procedure to estimate model parameters for predicting phenology. The main objective was to determine how well gene-based parameters performed relative to conventionally estimated parameters, testing the results with simulations of flowering time of diverse bread wheats grown over a wide range of wheat production environments.

MATERIALS AND METHODS

Data Sources

Data on crop management and phenology were obtained from the International Winter Wheat Performance Nurseries (IWWPN) from 1969 to 1981. This series of replicated trials was distributed to over 30 countries globally by the University of Nebraska and USDA-ARS with support from the U.S. Agency for International Development. In total, 168 cultivars were tested in the IWWPN. Each year, 30 cultivars or breeding lines were grown, and most cultivars were included for at least 2 yr. Trials were grown at 85 locations, which predominantly represented traditional winter wheat production areas but included spring wheat areas. Results of the IWWPN were published in 13 reports of the Research Bulletin of the Nebraska Agricultural Experiment Station (e.g., Kuhr et al., 1984). Phenotypic data and management summaries were digitized from the reports and stored in the International Wheat Information System (IWIS; Payne et al., 2002) at the International Maize and Wheat Improvement Centre (CIMMYT). Various minor errors such as in reporting of geographic coordinates of locations, year of sowing, and the reference date for time of flowering were corrected, so the data in IWIS differed slightly from that in the published reports. There was possible confusion over reporting of “time of flowering.” Although this term is logically associated with time of anthesis, most wheat programs only record heading date. Thus, the reported data may have referred to heading or anthesis. To permit meaningful comparisons across locations, all flowering data were re-expressed as days after sowing.

Although 168 cultivars were tested in the 13 cycles of the nurseries, only 29 cultivars were used (Table 1) due to limited availability of information on their genetic makeup. Alleles present at the *Ppd* and *Vrn* loci were determined through various sources. The online database (Martynov et al., 2006) provided

Table 1. Habit, origin, pedigree, mean observed days to anthesis, assumed genotypes, and estimated model parameters for cultivars used in the study.[†]

Cultivar	Habit	Origin	Pedigree	Days to anthesis		Locus				Model parameters	
				Mean	No. values [‡]	<i>Vrn-A1</i>	<i>Vrn-B1</i>	<i>Vrn-D1</i>	<i>Ppd-D1</i>	P1V	P1D
				d						d	%
INIA 66	S	Mexico	Lerma Rojo 64/Somora 64	181	15	1	1	0	1	18	29
Super X	S	Mexico	Penjamo-62(SIB)/Gabo-55	208	86	1	1	0	1	44	32
Lerma Rojo 64	S	Mexico	Yaqui-50//Norin-10/Brevor/3/ Lerma 52/4/2*Lerma Rojo	200	202	0	1	1	1	29	35
Bastion	S	Netherlands	Halle-12//Halle-35/Mara	220	64	0	1	0	1	48	48
Irnerio	S	Italy	Produttore/Manitoba	208	62	0	0	1	1	48	27
Aurora	W	USSR	Hard Federation/Cleveland//Sands	213	58	0	0	0	1	71	21
Balkan	W	Yugoslavia	Backa/Bezostaya-1// Mironovskaya-808/3/NS-433/4/ Skorospelka-35	214	37	0	0	0	1	62	21
Bezostaya 1	W	USSR	(S) Bezostaya-4	214	348	0	0	0	1	58	25
Biserka	W	Yugoslavia	Fortunato*2/(CI-13170)Redcoat	212	65	0	0	0	1	53	26
Dwarf Bezostaya	W	USSR	(S) Bezostaya	211	57	0	0	0	1	68	21
Jugoslavija	W	Yugoslavia	NS-646/Bezostaya 1//Aurora	214	38	0	0	0	1	62	31
Moslavka	W	Yugoslavia	ZG-3814-65/TP-114-1965-A// ZG-3814-65.Sanja	209	35	0	0	0	1	53	20
Odesskaya 51	W	USSR	Odesskaya-16/Bezostaya-1	218	72	0	0	0	1	65	18
Partizanka	W	Yugoslavia	Bezostaya-1/NS-116	215	71	0	0	0	1	62	15
Phoenix	W	Australia	WW-15*2/WW-80	205	38	0	0	0	1	55	17
San Pastore	W	Italy	Balilla/Villa Glori	194	42	0	0	0	1	50	17
Sanja	W	Yugoslavia	ZG-414-57/Leonardo	209	60	0	0	0	1	54	31
Sava	W	Yugoslavia	Fortunato*2/(CI-13170)Redcoat	211	58	0	0	0	1	37	37
Talent	W	France	Champlein/3/Thatcher/ Vilmorin-27//Fortunato	215	68	0	0	0	1	57	28
Zlatna dolina	W	Yugoslavia	ZG-414-57/Leonardo	209	65	0	0	0	1	62	24
Zlatoklasa	W	Yugoslavia	Sanja/TP-114-1965-A//Sanja	210	68	0	0	0	1	55	25
Bounty	W	England	Maris Ploughman/Durin	224	36	0	0	0	0	59	49
Cappelle Desprez	W	France	Vilmorin-27/Hybride du-Joncquois	216	23	0	0	0	0	68	31
Maris Huntsman	W	England	CI-12633/5*Cappelle- Desprez//Hybrid-46/Cappelle- D./3/2*Professeur-Marchal	226	67	0	0	0	0	59	48
Maris Mardler	W	England	Maris Ranger/Maris Durin//Maris Huntsman	219	38	0	0	0	0	60	39
Maris Nimrod	W	England	CI-12633/Yeoman//5*Cappelle/3/ Cappelle/Hybrid- 46/4/2*Professeur-Marchal	220	67	0	0	0	0	65	38
Maris Templar	W	England	CI-12633/5*Cappelle D.//Hei- nes-110/ Cappelle-D./3/Nord- Desprez/4/Viking	227	67	0	0	0	0	74	39
Mironovskaya 808	W	USSR	(T) Artemovka	219	72	0	0	0	0	72	34
Vakka	W	Finland	Varma/G-5-20-Kehra	222	60	0	0	0	0	57	42

[†]Habit is as reported in the International Winter Wheat Performance Nurseries (IWWPN); S, spring; W, winter. Origins are as reported in the original IWWPN reports and may not reflect current political boundaries. Loci were scored 0 or 1 if the recessive or dominant allele, respectively, is present. Cultivars are ordered from greatest number of dominant alleles within the *Vrn* and *Ppd* loci.

[‡]Number of locations and years used to calculate the mean for a cultivar.

data for many cultivars and also contained information on pedigrees and growth habit. Dencic (2001) and van Beem et al. (2005) contained additional data. Few evaluations for *Ppd* loci were found. Some daylength insensitive cultivars of European origin were classified by assuming that they contained the *Ppd-D1* locus (Worland, 1996; Worland et al., 1998).

Eighty-two locations of the IWWPN provided sufficient crop management data to allow simulating crop growth and development. Daily data for maximum and minimum tempera-

tures (required to simulate crop development) were obtained for 48 locations, representing 362 experiments (Table 2). Weather data sources included online databases (e.g., University of California, Agriculture and Natural Resources, 2005), individual researchers, and the Global Summary of the Day database (Lott, 1998). In several cases, the nearest weather station corresponded to an airport or other nonagricultural source, so elevation, distance from reported locations of experiments, and land use were examined using the Google Earth mapping tool

(Google, Inc., Mountain View, CA). No weather station data were used if the station was located more than 25 km from the experiment or if the pairs of locations differed greatly in elevation or land cover. No attempt was made to link locations to soil profile descriptions as simulations assumed no limitations of water and nutrients.

Phenotypic and management data were extracted from IWIS as spreadsheets and reformatted for use in the model using a structure similar to the standards of the International Consortium for Agricultural Systems Analysis (Hunt et al., 2001, 2006). Phenotypic data were available for days from sowing to flowering and maturity, grain yield, and other traits, but for this study, only flowering data were considered. The set of management data required to simulate an experiment thus was location (latitude), sowing date, and the list of cultivars grown in that trial.

Simulation Model

All simulations were conducted with the CSM-Cropsim-CERES-Wheat model Version 4.0.2.0 (Jones et al., 2003; Hoogenboom et al., 2004a), which incorporated features from Cropsim (Hunt and Pararajasingham, 1995) and CERES-Wheat (Ritchie, 1991; Ritchie et al., 1998). In CSM-Cropsim-CERES-Wheat, vernalization and photoperiod characteristics of different wheat cultivars or lines are specified in species, ecotype, and cultivar parameter files (Table 3). Developmental stages simulated include germination, seedling emergence, terminal spikelet initiation, anthesis, and physiological maturity. Species and ecotype parameters were held constant for all cultivars. The values were set as provided with the model except for the temperature response of vernalization and the critical long photoperiod, which are discussed below.

Rates of development vary with temperature and photoperiod. The model calculates the average daily temperature as the mean of the daily maximum and minimum temperatures, and all cardinal temperatures for development are based on average temperatures. In the absence of vernalization and photoperiod effects, development rates increase linearly at temperatures above a 0°C base temperature to an optimum of 26°C, above which development proceeds at a maximum rate. Occurrences of stages are simulated by integrating the effective temperatures over time, with a particular stage being reached when sufficient progress, quantified as accumulated effective temperature, has accrued.

In the presence of vernalization and/or photoperiod effects, the accumulated effective temperature is replaced by an equivalent temperature obtained by summing the products of daily effective temperature, a daily vernalization factor

Table 2. Locations providing data used for calibration or evaluation of the gene-based model parameters.

Country	Location	Latitude	Longitude	Elevation m	Experiments
Calibration					
Canada	Lethbridge	49.72	-112.80	909	4
Chile	Chillan	-35.47	-71.92	217	5
Chile	Temuco-Carillanca	-37.33	-72.58	332	6
Germany	Weihenstephan	48.40	11.73	467	12
Hungary	Martonvasar	47.35	18.82	150	11
Japan	Morioka Iwate	39.75	141.13	167	8
Mexico	Toluca	19.27	-99.85	2640	3
Netherlands	Wageningen	51.97	5.64	7	13
South Africa	Bethlehem	-27.83	28.30	1631	12
United States	Davis, CA	38.53	-121.75	15	8
United States	Brookston, IN	40.58	-86.93	183	4
United States	Hutchinson, KS	38.05	-97.92	460	5
United States	Ithaca, NY	42.45	-76.45	293	6
United States	Pullman, WA	46.70	-117.13	777	8
Evaluation					
Argentina	Balcarce	-36.25	-58.23	135	8
Argentina	Bordenave	-34.15	-63.02	212	13
Argentina	Pergamino	-32.12	-60.58	68	1
Austria	Vienna	48.20	16.75	147	11
Bulgaria	Dubrodja	43.72	28.18	236	11
Croatia	Zagreb	45.82	15.98	177	12
Czech Republic	Sedlec	50.23	14.50	300	6
Finland	Jokioinen	60.82	23.48	92	3
France	Orgerus	48.83	1.67	100	4
Germany	Monsheim	49.58	8.33	160	11
Hungary	Szeged	46.00	20.00	84	8
Iraq	Sulaimaniya	36.50	46.50	700	6
Iran	Hamadan	34.78	48.50	1200	8
Iran	Karaj	35.78	50.00	1300	9
Italy	Milano	45.22	9.42	68	11
Nepal	Kathmandu	27.67	85.33	1360	2
Romania	Fundulea	44.50	24.17	66	13
Russia	Krasnodar	45.00	38.92	31	8
South Korea	Suwon	36.32	126.98	37	12
Slovakia	Male Ripnany	48.48	17.98	172	7
Sweden	Svalof	55.58	13.10	50	7
Switzerland	Zurich	47.48	8.53	445	8
Turkey	Erzurum	39.97	41.33	1870	8
Turkey	Eskisehir	39.75	31.58	789	5
Ukraine	Mironovski	50.25	31.17	151	2
Ukraine	Odessa	46.45	30.70	42	1
United Kingdom	Cambridge	52.17	0.10	65	6
United States	Akron, CO	40.08	-103.67	1389	1
United States	Ft. Collins, CO	40.58	-105.17	1475	11
United States	Billings, MT	45.80	-108.53	923	6
United States	Mead, NE	41.17	-96.42	360	9
United States	Salisbury, NC	35.70	-80.62	251	9
United States	Stillwater, OK	36.12	-97.07	270	13
United States	Corvallis, OR	44.50	-123.50	68	7

Table 3. Major model parameters affecting time of anthesis and related traits as specified in CSM-Cropsim-CERES-Wheat species, ecotype and cultivar files.[†]

Parameter	Definition	Value(s)	Units
Species file			
P1DT	Optimal long photoperiod, above which there is no additional effect of photoperiod on development.	23 [‡]	h
P1VT	Vernalization type. Threshold value used to characterize response to vernalization.	50	day
P2(1)	Duration from terminal spikelet to jointing	80	°C day [§]
P4(1)	Relative duration from end of ear growth to anthesis	0.25	fraction of P4
PECM	Duration from germination to seedling emergence from a 1 cm planting depth.	10	°C day cm ⁻¹
PEG	Duration from sowing to germination in the absence of moisture stress.	10	°C day
TRDV1	Temperature response, development 1	(curve)	°C
TRDV2	Temperature response, development 2	(curve)	°C
TRVRN	Temperature response, vernalization	(curve) [†]	°C
TRLTH	Temperature response, lethal temperature hardening	(curve)	°C
WFGEU	Effect of soil water availability on seed germination	0.5	fraction of
Ecotype file			
P1	Duration from end of juvenile phase to double ridge formation	280	°C day
P2	Duration from double ridges to end of leaf growth	250	°C day
P3	Duration from end of leaf growth to end of spike growth	240	°C day
P4	Duration from end of spike growth to end of lag phase for grain filling	300	°C day
Cultivar file			
P1V	Days at optimum vernalizing temperature required to complete vernalization.	18–74	day
P1D	Percentage reduction in development rate in a photoperiod 10 h shorter than the optimum (P1DT) relative to the rate at the optimum photoperiod	15–48	%
PHINT	Phyllochron interval: the interval in thermal time between successive leaf tip appearances.	80	°C day

[†]Values for P1V and P1D are ranges among the 29 cultivars. All other values were constant across cultivars.

[‡]Values were modified from those provided in the official release of the model. See methods section for details.

[§]Degree day.

(ranging from 0 to 1), and a photoperiod factor (also 0 to 1). The required accumulations of equivalent temperature are model parameters (Table 3).

The cultivar-specific vernalization coefficient P1V specifies how many days of vernalization are required in order for flowering to occur, assuming that temperatures for vernalization are optimal. Vernalization was assumed to occur at temperatures from -4 to 15°C, with the maximum rate occurring between 0 and 3°C. These values differ from the cardinal temperatures provided in the official release of the model (Hoogenboom et al., 2004b), which had cardinal temperatures of -5, 0, 7, and 15°C. This change was introduced because initial tests showed that the original cardinal temperatures required allowing values of P1V as high as 95 d, whereas most studies suggest that complete vernalization occurs in less than 60 d (e.g., Davidson et al., 1985). The model also allows for devernialization to occur when less than 10 d of progress toward vernalization have accumulated, and the maximum temperature exceeds 30°C. The daily vernalization rates are accumulated to indicate vernalization status, this value being used to calculate the vernalization factor. The daily photoperiod factor is calculated using a curvilinear response with P1D and P1DT as parameters, and photoperiod as the environmental variable.

Cultivar sensitivity to photoperiod is mainly determined by the parameter P1D, which specifies the reduction in developmental rate in a photoperiod 10 h shorter than the critical long photoperiod, P1DT. The photoperiod is calculated using a criterion of the sun being 6° below the horizon. P1DT was

increased from 20 to 23 h for all cultivars to increase responsiveness of the model to the long photoperiods that occurred at high latitude sites.

Model Calibration

Fourteen locations, representing 105 experiments, were used to calibrate the model (Table 2). Each cultivar was calibrated independently starting from initial values of P1V of 60 d and P1D of 50%. These values were alternately modified, and simulations were run over the calibration datasets. Goodness of fit was judged by comparing means of observed vs. simulated values of flowering date, r^2 values, and RMSE. All other species, ecotype, and cultivar parameters (Table 3) were held constant.

Estimation of Genetic Effects

Genetic effects on P1V and P1D were estimated using the calibrated P1V and P1D values, and a linear regression approach similar to the procedure of White and Hoogenboom (1996) in which each locus is coded with a value of 1 for dominant and 0 for recessive. For vernalization, for which evidence from Halloran (1967) suggests that the *Vrn-1* loci have roughly additive effects (White, 2006), the relationship was established was $P1V = 60.0 - 14.5N_{Vrn}$, where N_{Vrn} is the total number of *Vrn-1* loci present (dominant) in a given cultivar ($r^2 = 0.57$, with $P < 0.01$; RMSE = 8.3 d).

For photoperiodism, although three *Ppd* loci have been identified in wheat, sufficient data were only available to estimate the effect of the *Ppd-D1* locus. The resulting linear model

for photoperiod was $P1D = 39.9 - 13.9Ppd-D1$, where $Ppd-D1$ was coded with a value of 1 for wheat cultivars that possessed the dominant allele, and a value of 0 for the recessive allele. This regression gave an r^2 value of 0.42 ($P < 0.01$) and a RMSE of 7.5%. The reduction in sensitivity indicated in the equation agrees with expectation that dominant alleles of Ppd reduce photoperiod sensitivity (e.g., Worland, 1996).

Model Evaluation

Phenotypic data from 34 locations, representing 257 experiments were used to evaluate model outputs. Since locations used for calibration were independent of those used for evaluation, the evaluation results were expected to have less bias than if data from the same location were used both for calibration and evaluation. Initial evaluations used bivariate regressions of observed vs. simulated values. Multiple regressions were then used to assess the relative predictive capability of cultivar coefficients estimated directly or indirectly using knowledge of the genetic make-up of the cultivars and the equations linking genetic composition to the P1V and P1D coefficients, as in the approach of White et al. (2007). For the assessments, it was of interest to have a baseline value that would represent a generic cultivar. Thus, values of P1V and P1D were estimated for a hypothetical, generic cultivar by calculating mean values for the 29 cultivars. In regressions comparing predictions based on the three types of estimates of P1V and P1D, predicted values were considered fixed effects and were analyzed using the GLM procedure the SAS package (version 9.1, SAS Institute, Cary, NC). For analyses examining remaining sources of variation, locations, years, and cultivars were considered random effects, and variance components were estimated using the restricted maximum likelihood method as implemented for the MIXED procedure of SAS.

RESULTS

Simulations using the conventionally estimated coefficients predicted 98% of variation in days to flowering for

Table 4. Summary of comparisons of observed vs. simulated values of days to flowering for conventionally estimated model coefficients, gene-based estimates, and coefficients for a single generic cultivar.[†]

	Mean	r^2	Slope	Intercept	RMSE
	d				d
Calibration ($n = 540$)[‡]					
Observed values	208				
Conventional	209	0.98	0.95 (0.01)	10.1 (1.4)	6.6
Gene-based	208	0.96	0.92 (0.01)	16.0 (1.7)	8.6
Generic cultivar	208	0.95	0.95 (0.01)	10.0 (2.0)	9.7
Evaluation ($n = 1499$)					
Observed values	214				
Conventional	213	0.92	1.00 (0.01)	0.6 (1.6)	9.0
Generic cultivar	212	0.90	0.99 (0.01)	5.2 (1.8)	9.9
Mean values	214	0.89	1.02 (0.01)	-2.7 (2.0)	10.4

[†]Values of r^2 are significant at the 0.001 probability level. Values in parentheses after slopes and intercepts are standard errors.

[‡] n = number of pairs of observed vs. simulated data with no missing values.

the calibration dataset and 92% for the evaluation dataset (Table 4 and Fig. 1). The calibration dataset also showed a RMSE of 6.6 vs. 9.0 d for the evaluation data, agreeing with the expected loss of accuracy when dealing with an independent set of data. For single nurseries (years of distribution) of the IW WP N and across all locations, the difference in mean days to flower for the earliest and latest cultivars was typically 18 to 21 d.

Simulations with the gene-based coefficients gave r^2 values of 0.96 for the calibration and 0.90 for the evaluation data, with respective RMSE of 8.6 and 9.9 d. While the results with the gene-based model suggested only moderate loss of accuracy as compared to conventional coefficients, simulations with single values of P1V and P1D, representing the generic cultivar, also gave r^2 values of 0.95 for the calibration data and 0.89 for the evaluation data and RMSE values of 9.7 and 10.4 d, respectively (Table 4).

Regression analysis permitted quantifying whether modeling based on genetic information and relationships linking genetic composition to model coefficients resulted in a significant loss in accuracy. Both the gene-based and conventional coefficients explained large portions ($P < 0.001$) of the variation in the residuals from the regression for simulations with the generic cultivar (Table 5). The remaining variation was estimated as the total sums of squares (SS) less SS attributed to the generic cultivar in Table 5. Thus, for the calibration dataset, the gene-based model explained 29% of remaining variation, and the conventional model, 54%. For the evaluation dataset, the gene-based model explained 17% of remaining variation, and the conventional model, 27%. Comparing the residuals from the regression for simulations with the generic cultivar with the simulations using the other coefficients revealed that this explanatory ability was related to the difference in phenology between spring and winter

Table 5. Analyses of variance for linear regressions examining ability of simulations to account for observed variation in phenology, considering both calibration and evaluation datasets.[†]

Source	df	Sums of squares (SS)	% SS	F-value	RMSE
					d
Calibration					
Generic cultivar	1	943,086	94.8	21,474	6.6
Gene-based	1	14,812	1.5	337	
Conventional	1	12,926	1.3	294	
Residual	536	23,540	2.4		
Evaluation					
Generic cultivar	1	1,371,807	89.5	17,269	8.9
Gene-based	1	27,452	1.8	346	
Conventional	1	16,442	1.1	207	
Residual	1481	117,649	7.7		

[†]The ANOVAs are for sequential entry of effects using values of P1V and P1D for the generic cultivar, gene-based estimates, and conventionally estimated values. F-values are significant at the 0.001 probability level.

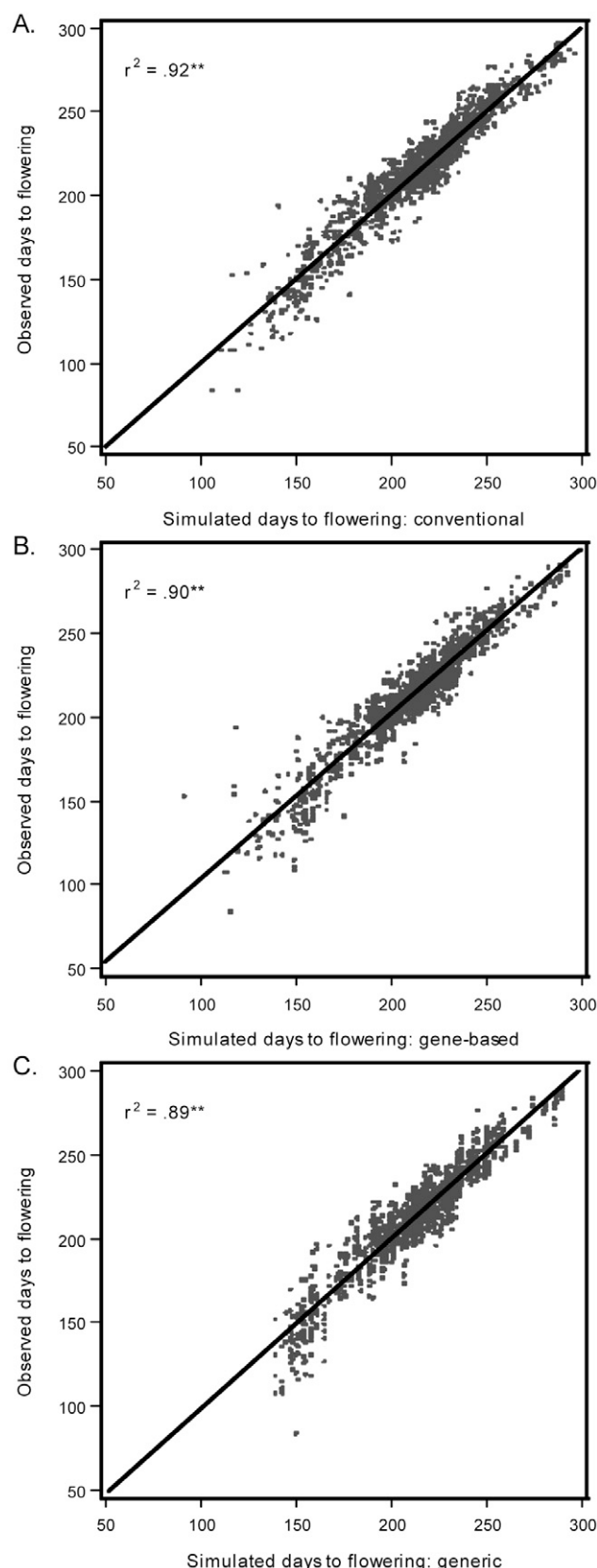


Figure 1. Comparisons of observed vs. simulated days to flowering for the evaluation datasets using different values of the coefficients P1V and P1D for the simulations: (A) conventionally estimated coefficients; (B) gene-based coefficients; (C) coefficients for a single generic cultivar (estimated as means of the conventional estimates). Fitted lines are for linear regressions with slopes and intercepts as given in Table 4.

wheats and the variation within the spring cultivars (Fig. 2, evaluation dataset only).

Complementary information about the performance of the gene-based coefficients was obtained with the data for residuals by first accounting for simulations using the gene-based coefficients, and then testing for effects of location, year within location, cultivar, and simulations using the conventionally determined coefficients (Table 6). When used alone, the gene-based coefficients explained over 90% of the total variation in the evaluation dataset (Table 4), but large portions of variation in the residuals were still associable with effects of location, year, cultivar, and the conventional model (Table 6).

DISCUSSION

The predictive capability of the model using gene-based coefficients confirmed the potential for predicting wheat phenology by combining physiological descriptions with genetic data. For the evaluation dataset, the gene-based model explained 63% of the variation that would be explained by conventionally determined model coefficients (SS of 14,812 for gene-based alone vs. SS of 27,738 for the conventional model), once the generic cultivar effect was considered (Table 5). Furthermore, when flowering date was simulated with the gene-based approach, only 5% of the total variation was unexplained in the calibration dataset and 10% in the evaluation set (Table 4).

The comparisons of conventional and gene-based simulations with residuals from the generic cultivar (Fig. 2) illustrated the impact of the limited availability of data on specific loci. There was no additional predictive power within the winter wheats, where only two genotypes (*Ppd-D1* dominant or recessive) were distinguishable with the available genetic data.

The results arguably present an optimistic view of the potential of gene-based wheat modeling because the set of genotypes and locations is more diverse than routinely dealt with in wheat breeding nurseries. Spring and winter wheats are seldom evaluated together, and experimental sites are often grouped according to similarities in growing environment and germplasm types grown. Nonetheless, first accounting for effects of the generic model should have fully compensated for the bias from considering a wide range of environments.

Conversely, the results may present a conservative impression of the utility of gene-based approaches due to constraints on the accuracy and completeness of the genotypic, phenotypic, and environmental data and of the model calibration. Our ability to ascribe genotypes for *Vrn* and *Ppd* loci should improve with rapid progress in sequencing the loci and understanding how they interact (Yan et al., 2003; Sherman et al., 2004; Turner et al., 2005). Furthermore, the evidence for multiple alleles of the *Vrn-1* loci (Košner and Pánková, 1998; Sherman et al.,

2004; Yan et al., 2004) and for additional *Vrn* loci (Yan et al., 2006) offers promise for improved discrimination of vernalization response. Similarly, variation for earliness per se (*Eps*) loci (Scarath and Law, 1984; Appendino and Slafer, 2003; Tóth et al., 2003; van Beem et al., 2005) were not considered due to insufficient information on variation in *Eps* loci among cultivars. Incorporating effects of the *Eps* loci not only should improve overall performance of the model, but should allow for more accurate characterization of effects of the *Vrn* and *Ppd* loci.

Evaluations of time to flowering likely were subject to observer bias. In wheat research, “flowering” is most often equated to time of spike emergence, which is also termed “ear emergence” or “heading.” However, it can also refer to time of anthesis. Regardless of the reference stage, there also may have been observer bias as to what portion of plants (or tillers) must have individually attained a given stage before a field plot was scored as having reached that stage. Weather data were also problematic. Although considerable effort was made to match reported experiment locations to weather stations, lack of weather data from actual experiment sites undoubtedly introduced additional error. Unlike most assessments of ecophysiological models, the evaluation data were fully independent of the calibration data since the two sets represented geographically separate sets of locations, which furthermore were managed by different research teams at each location.

There is no question of the long-term value of understanding the control of flowering at the level of gene expression and molecular signals, and recent progress is remarkable (Trevaskis et al., 2007). Our positive results emphasize, however, the immediate utility of combining existing ecophysiological models with genetic data. In the case of wheat, progress in interpreting genotype by environment interactions appears readily obtainable by improving characterization of wheat lines for variation in the *Vrn* and *Ppd* loci. Such an approach can benefit from information emerging from molecular studies, but the characterizations should emphasize robust description of ecotypic variation over detailed understanding of molecular mechanisms. Progress in quantitative modeling of phenology will likely reveal areas where research on specific processes will have high payoff. Ecophysiological models thus may provide a valuable tool for priority setting and hypothesis testing in plant biology, echoing arguments of R.S. Loomis and colleagues almost 30 yr ago for appropriate use of models (Loomis et al., 1979).

The strength of the results rest in large part on the availability of the large IWGPN dataset, currently maintained by CIMMYT. Although conducted over 25 yr ago, the experiments provide a unique resource for examining responses of genotype to environment. Combining findings from genomics with the integrative power of ecophysiological models holds potential for increasing our ability to

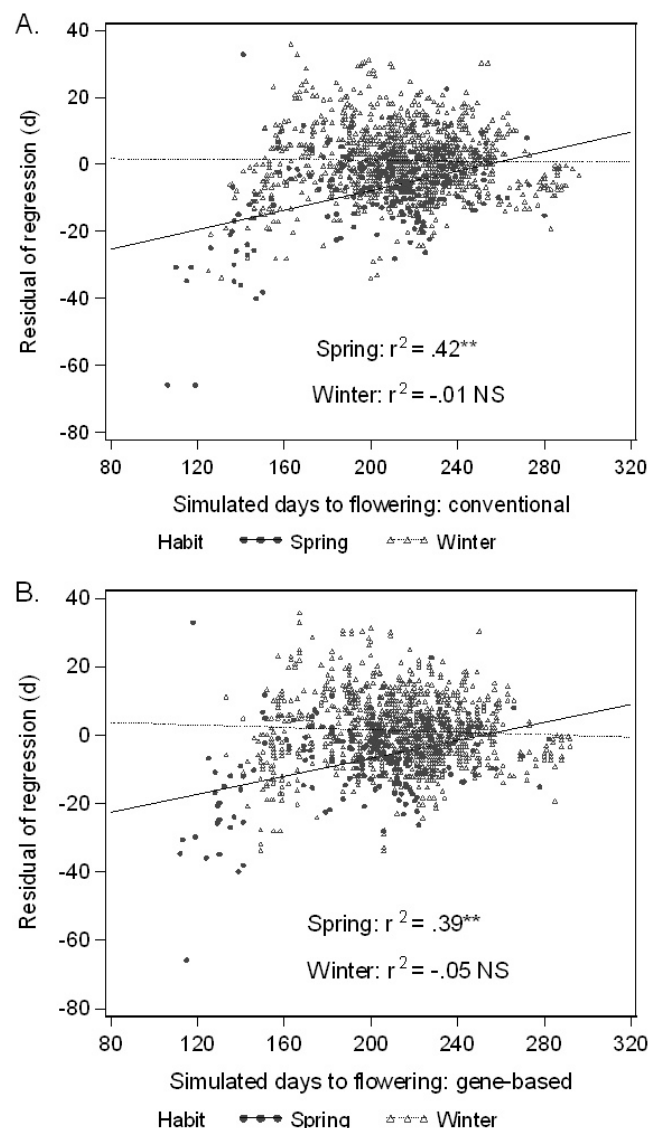


Figure 2. Comparisons of residuals of regression with the generic cultivar vs. simulated days to flowering for the evaluation datasets using different values of the coefficients P1V and P1D for the simulations. Fitted lines are for linear regressions estimated separately for spring or winter types: (A) conventionally estimated coefficients; (B) gene-based coefficients.

predict crop performance as a function of genotypes, management and expected environmental conditions.

CONCLUSIONS

Once expected large effects of location and year were accounted for, simulations using the gene-based approach accounted for 29 and 17% of variation in time to flowering, for the calibration and validation datasets, respectively. In comparison, conventional coefficients explained 54 and 27% of variation for the two datasets. Additional information on genetic makeup of cultivars should further improve the predictive power of the gene-based simulations. This might involve either refinements of linear estimates of gene effects or more mechanistic representations based on emerging information on gene action. Although

Table 6. Restricted maximum likelihood method analysis examining how location, year, cultivar, and model predictions from conventionally estimated coefficients account for residual variation in observed time to flower after allowing for model predictions using gene-based coefficients (regression equations given in Table 4).[†]

Parameter	Calibration dataset			Evaluation dataset		
	Variance component	SE	Z-value	Variance component	SE	Z-value
Location	17.83	8.91	2.00*	171.28	49.68	3.45***
Year (location)	14.89	2.87	5.18***	35.60	3.96	8.98***
Cultivar	1.85	0.80	2.33**	8.56	2.51	3.42***
Residual	15.90	1.14	13.98***	14.25	0.60	23.69***

Fixed effects	Value	SE	df	t	Value	SE	df	t
Intercept	21.36	4.02	13	5.3***	70.99	5.06	28	14.0***
Conventional	0.89	0.02	406	49.7***	0.68	0.02	1210	32.8***

*Significant at the $P = 0.05$ probability level.

**Significant at the $P = 0.01$ probability level.

***Significant at the $P = 0.0001$ probability level.

[†]Location, year within location, and cultivar were considered random effects, and the model predictions as fixed effects.

the focus was on phenology, a gene-based approach should be extensible to traits that more directly affect yield or grain quality. Application to other traits, however, again requires improved genetic characterization of wheat lines. The basic approach also appears readily applicable to other wheat models as well as to other crops.

Acknowledgments

The authors would like to recognize the efforts of the international wheat research community in executing the IWHPN. We also thank the numerous scientists who provided weather historical data or otherwise clarified issues relating to the IWHPN data. Portions of this work were supported through Deutsche Forschungsgemeinschaft (DFG), Grant No. 1070/1, International Research Training Group “Sustainable Resource Use in North China.”

References

- Appendino, M.L., and G.A. Slafer. 2003. Earliness per se and its dependence upon temperature in diploid wheat lines differing in the major gene *Eps-Am1* alleles. *J. Agric. Sci.* 141:149–154.
- Asseng, S., N.C. Turner, J.D. Ray, and B.A. Keating. 2002. A simulation analysis that predicts the influence of physiological traits on the potential yield of wheat. *Eur. J. Agron.* 17:123–141.
- Baenziger, P.S., G.S. McMaster, W.W. Wilhelm, A. Weiss, and C.J. Hays. 2004. Putting genes into genetic coefficients. *Field Crops Res.* 90:133–143.
- Davidson, J.L., K.R. Christian, D.B. Jones, and P.M. Bremner. 1985. Responses of wheat to vernalization and photoperiod. *Aust. J. Agric. Res.* 36:347–359.
- Dencic, S. 2001. Yugoslav wheat pool. p. 377–404. *In* A.P. Bonjean and W.J. Angus (ed.) *The world wheat book*. Lavoisier Publishing, Paris.
- Edmeades, G.O., G.S. McMaster, J.W. White, and H. Campos. 2004. Genomics and the physiologist: Bridging the gap between genes and crop response. *Field Crops Res.* 90:5–18.
- Flood, R.G., and G.M. Halloran. 1986. Genetics and physiology of vernalization response in wheat. *Adv. Agron.* 39:87–125.
- Grant, R.F., B.A. Kimball, T.J. Brooks, G.W. Wall, P.J. Pinter, Jr., D.J. Hunsaker, F.J. Adamsen, R.L. Lamorte, S.W. Leavitt, T.L. Thompson, and A.D. Matthias. 2001. Modeling interactions among carbon dioxide, nitrogen, and climate on energy exchange of wheat in a free air carbon dioxide experiment. *Agron. J.* 93:638–649.
- Hammer, G.L., T.R. Sinclair, S.C. Chapman, and E. Van Oosterom. 2004. On systems thinking, systems biology, and the in silico plant. *Plant Physiol.* 134:909–911.
- Halloran, G.M. 1967. Gene dosage and vernalization response in homoeologous group 5 of *Triticum aestivum*. *Genetics* 57:401–407.
- Hoogenboom, G., J.W. Jones, P.W. Wilkens, C.H. Porter, W.D. Batchelor, L.A. Hunt, K.J. Boote, U. Singh, O. Uryasev, W.T. Bowen, A.J. Gijssman, A. du Toit, J.W. White, and G.Y. Tsuji. 2004a. Decision support system for agrotechnology transfer version 4.0. [CD-ROM]. Univ. of Hawaii, Honolulu.
- Hoogenboom, G., and J.W. White. 2003. Improving physiological assumptions of simulation models by using gene-based approaches. *Agron. J.* 95:82–89.
- Hoogenboom, G., J.W. White, J. Acosta-Gallegos, R. Gaudiel, J.R. Myers, and M.J. Silbernagel. 1997. Evaluation of a crop simulation model that incorporates gene action. *Agron. J.* 89:613–620.
- Hoogenboom, G., J.W. White, and C.D. Messina. 2004b. From genome to crop: Integration through simulation modeling. *Field Crops Res.* 90:145–163.
- Hunt, L.A., G.H. Hoogenboom, J.W. Jones, and J.W. White. 2006. ICASA version 1.0 data standards for agricultural research and decision support. Available at www.icasa.net/standards/index.html (verified 23 Jan. 2008). Int. Consortium for Agricultural Systems Applications, Honolulu, HI.
- Hunt, L.A., and S. Pararajasingham. 1995. Cropsim-Wheat: A model describing the growth and development of wheat. *Can. J. Plant Sci.* 75:619–632.
- Hunt, L.A., J.W. White, and G. Hoogenboom. 2001. Agronomic data: Advances in documentation and protocols for exchange and use. *Agric. Syst.* 70:477–492.
- Iqbal, M., A. Navabi, R.-C. Yang, D.F. Salmon, and D. Spaner. 2007. The effect of vernalization genes on earliness and related agronomic traits of spring wheat in northern growing regions. *Crop Sci.* 47:1031–1039.
- Jamieson, P.D., J.R. Porter, J. Goudriaan, J.T. Ritchie, H. van Keulen, and W. Stol. 1998. A comparison of the models

- AFRCWHEAT2, CERES-Wheat, Sirius, SUCROS2, and SWHEAT with measurements from wheat grown under drought. *Field Crops Res.* 55:23–44.
- Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijssman, and J.T. Ritchie. 2003. The DSSAT cropping system model. *Eur. J. Agron.* 18:235–265.
- Košner, J., and K. Pánková. 1998. The detection of allelic variants at the recessive *vrn* loci of winter wheat. *Euphytica* 101:9–16.
- Kuhr, S.L., C.J. Peterson, V.A. Johnson, P.J. Mattern, and J.W. Schmidt. 1984. Results of the thirteenth International Winter Wheat Performance Nursery grown in 1981. *Nebraska Agric. Exp. Stn. Res. Bull.* No 305. Univ. of Nebraska, Lincoln.
- Laurie, D.A., S. Griffiths, and V. Christodoulou. 2004. Comparative genetic approaches to the study of control of flowering time in temperate cereals. *Field Crops Res.* 90:87–99.
- Loomis, R.S., R. Rabbinge, and E. Ng. 1979. Explanatory models in crop physiology. *Annu. Rev. Plant Physiol.* 30:339–367.
- Lott, N. 1998. Global surface summary of day. Available at www.ncdc.noaa.gov/cgi-bin/res40.pl?page=gsod.html (verified 23 Jan. 2008). National Climatic Data Center, Asheville, NC.
- Martynov, S.P., T.V. Dobrotvorskaya, I. Hon, and I. Faberova. 2006. Wheat pedigree and identified alleles of genes. Available at genbank.vurv.cz/wheat/pedigree/default.htm (verified 23 Jan. 2008). Crop Research Institute, Prague.
- Messina, C.D., J.W. Jones, K.J. Boote, and C.E. Vallejos. 2006. A gene-based model to simulate soybean development and yield responses to environment. *Crop Sci.* 46:456–466.
- Minorsky, P.V. 2003. Achieving the in silico plant: Systems biology and the future of plant biological research. *Plant Physiol.* 132:404–409.
- Nakagawa, H., J. Yamagishi, N. Miyamoto, M. Motoyama, M. Yano, and K. Nemoto. 2005. Flowering response of rice to photoperiod and temperature: A QTL analysis using a phenological model. *Theor. Appl. Genet.* 110:778–786.
- Payne, T.S., B. Skovmand, C.G. Lopez, E. Brandon, and A. McNab. 2002. The international wheat information system (IWIS), version 4, 2001. [CD-ROM]. CIMMYT, Mexico, DF.
- Reymond, M., B. Muller, A. Leonardi, A. Charcosset, and F. Tardieu. 2003. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiol.* 131:664–675.
- Ritchie, J.T. 1991. Wheat phasic development. p. 31–54. *In* J. Hanks and J.T. Ritchie (ed.) *Modeling plant and soil systems*. ASA, CSSA, and SSSA, Madison, WI.
- Ritchie, J.T., U. Singh, D.C. Godwin, and W.T. Bowen. 1998. Cereal growth, development and yield. p. 79–98. *In* G.Y. Tsuji, G. Hoogenboom, and P.K. Thornton (ed.) *Understanding options for agricultural production*. Kluwer Academic, Dordrecht, the Netherlands.
- Scarath, R., and C.N. Law. 1984. The control of day-length response in wheat by the group 2 chromosomes. *Z. Pflanzenzucht.* 92:140–150.
- Sherman, J.D., L. Yan, L. Talbert, and J. Dubcovsky. 2004. A PCR marker for growth habit in common wheat based on allelic variation at the VRN-A1 Gene. *Crop Sci.* 44:1832–1838.
- Struik, P.C., X. Yin, and P.D. Visser. 2005. Complex quality traits: Now time to model. *Trends Plant Sci.* 10:513–516.
- Tardieu, F. 2003. Virtual plants: Modeling as a tool for the genomics of tolerance to water deficit. *Trends Plant Sci.* 8:9–14.
- Tóth, B., G. Galiba, E. Feher, J. Sutka, and J.W. Snape. 2003. Mapping genes affecting flowering time and frost resistance on chromosome 5B of wheat. *Theor. Appl. Genet.* 107:509–514.
- Trevaskis, B., M.H. Hemming, E.S. Dennis, and W.J. Peacock. 2007. The molecular basis of vernalization-induced flowering in cereals. *Trends Plant Sci.* 12:352–357.
- Turner, A., J. Beales, S. Faure, R.P. Dunford, and D.A. Laurie. 2005. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* 310:1031–1034.
- University of California, Agriculture and Natural Resources. 2005. UC IPM Online: Weather data and products. Available at www.ipm.ucdavis.edu/WEATHER/wxretrieve.html (verified 23 Jan. 2008). Statewide IPM Program, Agriculture and Natural Resources, Univ. California, Davis.
- van Beem, J., V. Mohler, R. Lukman, M. van Ginkel, M.C.J. William, and A.J. Worland. 2005. Analysis of genetic factors influencing the developmental rate of globally important CIMMYT wheat cultivars. *Crop Sci.* 45:2113–2119.
- Welch, S.M., J.L. Roe, and Z. Dong. 2003. A genetic neural network model of flowering time control in *Arabidopsis thaliana*. *Agron. J.* 95:71–81.
- White, J.W. 2006. From genome to wheat: Emerging opportunities for modelling wheat growth and development. *Eur. J. Agron.* 25:79–88.
- White, J.W., K.J. Boote, G. Hoogenboom, and P.G. Jones. 2007. Regression-based evaluation of ecophysiological models. *Agron. J.* 99:419–427.
- White, J.W., and G. Hoogenboom. 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agron. J.* 88:416–422.
- White, J.W., and G. Hoogenboom. 2003. Gene-based approaches to crop simulation: Past experiences and future opportunities. *Agron. J.* 95:52–64.
- Wollenweber, B., J.R. Porter, and T. Lubberstedt. 2005. Need for multidisciplinary research towards a second green revolution. *Curr. Opin. Plant Biol.* 8:337–341.
- Worland, A.J. 1996. The influence of flowering time genes on environmental adaptability in European wheats. *Euphytica* 89:49–57.
- Worland, A.J., A. Börner, V. Korzun, W.M. Li, S. Petrovic, and E.J. Sayers. 1998. The influence of photoperiod genes on the adaptability of European winter wheats. *Euphytica* 100:385–394.
- Yan, L., D. Fu, C. Li, A. Blechl, G. Tranquilli, M. Bonafede, A. Sanchez, M. Valarik, S. Yasuda, and J. Dubcovsky. 2006. The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proc. Natl. Acad. Sci. USA* 103:19581–19586.
- Yan, L., M. Helguera, K. Kato, S. Fukuyama, J. Sherman, and J. Dubcovsky. 2004. Allelic variation at the VRN-1 promoter region in polyploid wheat. *Theor. Appl. Genet.* 109:1677–1686.
- Yan, L., A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima, and J. Dubcovsky. 2003. Positional cloning of the wheat vernalization gene VRN1. *Proc. Natl. Acad. Sci. USA* 100:6263–6268.
- Yin, X., S.D. Chasalow, C.J. Dourleijn, P. Stam, and M.J. Kropff. 2000. Coupling estimated effects of QTLs for physiological traits to a crop growth model: Predicting yield variation among recombinant inbred lines in barley. *Heredity* 85:539–549.
- Yin, X., P.C. Struik, and M.J. Kropff. 2004. Role of crop physiology in predicting gene-to-phenotype relationships. *Trends Plant Sci.* 9:426–432.